

UNITED STATES DISTRICT COURT
SOUTHERN DISTRICT OF NEW YORK

.....X
MELISSA KAYE,

PLAINTIFF,

CASE NO.
18-CV-12137 (JPO)

-against-

NEW YORK CITY HEALTH AND HOSPITALS CORPORATION;
ELIZABETH FORD; PATRICIA ("PATSY") YANG; ABHISHEK JAIN;
and JONATHAN WANGEL (said names being fictitious, the
persons intended being those who aided and abetted the
unlawful conduct of the named Defendants

PARTIES' PROPOSED
PROTOCOL RELATING TO
THE PRODUCTION OF
ELECTRONICALLY STORED
INFORMATION

DEFENDANTS.
.....X

A. Scope

1. **General:** The procedures and protocols enumerated herein govern the production and exchange of electronically stored information ("ESI") between: the New York City Health and Hospitals Corporation, Elizabeth Ford, Patricia Yang, Abhishek Jain, and Jonathan Wangel (collectively referenced henceforth as "Defendants" or "Responding Party"); and Melissa Kaye, M.D. (referenced henceforth as "Plaintiff" or "Requesting Party"). The Parties have discussed the methodologies or protocols for the search and review of ESI. Pursuant to the tenets enumerated in Fed. R. Civ. P. 26, 34 and the Sedona Conference, the parties have sought to engage each other in good faith. Therefore, this protocol provides a general framework for the production of ESI on a rolling basis.

2. **Limitations and No-Waiver:** Nothing in this protocol will be interpreted to require the disclosure of documents or information protected from disclosure by the attorney-client privilege, work-product doctrine or any other applicable privilege or immunity. This agreement thereby seeks to capture the scope, methodology and obligations of the parties to produce

relevant, responsive and non-privileged ESI in a transparent and efficient fashion. In addition to this agreement, the parties have drafted a proposed Order for consideration pursuant to Fed.

R. Evid. 502. All documents produced in accordance with this protocol: are fully protected and covered by the Parties' confidentiality and claw back agreements; and the produced documents are also protected by any orders of the Court that may be issued to protect them.

3. Governance: The procedures described herein govern the use of predictive coding to assist in the production of documents by the Parties to this Action. In this Order, predictive coding shall mean and refer to a process for selecting and ranking the collection of documents using a computerized system that incorporates the decisions that the lawyers have made on a smaller set of documents and then applying those decisions to the remaining universe of documents.

4. Alternative Methodologies: Nothing in this Stipulation shall prevent the Responding Party from using other search, review or coding methodologies in addition to, or in place of, predictive coding to help identify documents that are responsive to the document requests from the Requesting Party. Alternatively, nothing in this agreement shall preclude the Plaintiff from requesting the utilization of keyword searches or other EDiscovery methodologies should it be determined that initial productions yielded inadequate or non-transparent responses.

5. Definitions

- a. "Precision" means the number of documents identified by the Predictive Coding Software as responsive that are actually responsive expressed as a percentage of the total number of documents identified as responsive.

- b. “Recall” means the number of responsive documents identified by the Predictive Coding Software expressed as a percentage of the total number of responsive documents in the Document Universe.
- c. “Statistically Valid Sample” means a sample of sufficient size and composition to permit statistical extrapolation with a margin of error of +/-2% at the 95% confidence level.

D. Search Methodology

1. Initial Disclosure of Information

Prior to the commencement of any review using predictive coding, the Responding Party shall disclose to Requesting Party in writing its intention to use predictive coding and the following information:

- a) The name, publisher, version number, and technical description of the software it intends to use to perform the predictive coding (the “Predictive Coding Software”);
- b) The name and qualifications of the person who will oversee the implementation of the predictive coding process (the “Subject Matter Expert”);
- c) A description of the documents to be subjected to predictive coding (the “Document Universe”), including but not limited to:
 - i) Custodian/Source;
 - ii) Data types (e.g. email, electronic documents, etc.);

- iii) Date range, search terms, and any other criteria
(the “Culling Criteria”) used to reduce the
volume of the Document Universe, and the
number of documents removed by the Culling
Criteria (the “Excluded Documents”)
- iv) The number of documents in the Document
Universe, in total and for each
Custodian/Source.

2. Responsiveness Categories

- a) The Defendants shall reveal the following about the responsiveness categories into which the Document Universe is to be categorized. The Parties shall meet and confer to address any questions or disputes about:
 - the selection of the Predictive Coding Software;
 - the Subject Matter Expert;
 - the Document Universe;
 - the Culling Criteria; and
 - the Responsiveness Categories.
- b) The Responding Party shall make its Subject Matter Expert reasonably available to address questions about the technical operation of the Predictive Coding Software.

3. Designation of Methodology; Reservation to Request Alternate Forms of Review and Defendants’ Capabilities

a) Initial Designation of Search Methodology: The parties have initially agreed to the use of **Predictive Coding/Continuous Active Learning (CAL) TAR 2.0**. However, Plaintiff reserves the right to request the use of TAR 1.0 or keyword searches should Defendants' production of documents prove inadequate. Based on conversations with Defendants, they have the capacity to engage in both CAL/TAR 2.0 and keyword searches/TAR 1.0 review.

b) Exhaustion of Searches and Court Intervention Before Depositions: If Defendants fail to engage in the transparent and good faith production under CAL, Plaintiff will first attempt to meet and confer with Defendants to discuss alternate review options. If such discussions prove futile, Plaintiff will then seek the Court's intervention. The parties will engage in this exercise prior to the taking of any depositions in this matter. Resolution of methodology and the adequacy of any search measures shall be established and if necessary confirmed by the Court prior to the taking of any depositions.

c. Overview of Predictive Coding Process

Defendants will initially utilize their CAL based platform to conduct ESI in this matter. The search process will begin with Defendants' attorneys developing an understanding of the entire universe of ESI from the relevant custodians, specified time periods and subject matters. The objectives are to determine Precision and Recall by watching the batches of documents culled for each search. During the course of this process, Defendants will identify a small number of documents that they will share with Plaintiff's counsel. The

parties will agree to this control group of documents before the commencement of any CAL based searches

The initial control group of documents, shall be representative of the categories to be reviewed and coded for the following:

- i) relevance, privilege and or issue-relation; and or**
- ii) alternatively short query, topic description or relevant document.**

d. Issue Tags

A) Identification of Issues: The parties agree that, to the extent applicable, as part of the initial training for the CAL searches described above, as well as during the iterative process, all documents categorized as relevant and not privileged, shall also be coded with one of the following agreed upon issue tags:

- i) FMLA
- ii) Pay Parity
- iii) Discrimination
- iv) Complaints/HR/Corporate Compliance
- v) Shift Change
- vi) Retaliation

B) Step by Step Protocol

Step 1: Generation of Control Set

- (i) Ideally, this exercise will generate a small set of documents that will be devoid of false representations and should be topically representative of the claims raised in the case. Calculating precision recall shall take place throughout the course of

this litigation and with the input of both parties. Since the topic descriptions, relevant documents and or queries may evolve due to motion practice or other issues raised in subsequently filed pleadings, the parties anticipate that review will be ongoing and collaborative throughout discovery.

- (ii) Precision Recall will assist the attorneys' identification of probative documents for each category to be reviewed and coded. Plaintiff's counsel will be provided with the preliminary results of Defendants' initial document selection for CAL and will be invited to share their own proposed words or topics for searches. Thereafter, Plaintiff's counsel will be provided with the non-privileged analytics from the first search. At this juncture, the Parties will discuss projected richness, confidence levels and review rates. Plaintiff's counsel will review the documents produced and will promptly provide defense counsel with their own evaluation of which coding should apply. To the extent the parties disagree regarding the coding of a particular document, they will meet and confer in an effort to resolve the dispute prior to contacting the Court for resolution. The irrelevant documents so produced shall be promptly returned after review and analysis by Plaintiff's counsel and/or resolution of any disputes by the Court.
- (iii) This initial set of control documents will be used to begin the Predictive Coding process. Each control set of documents will be applied to its relevant category and will start the "training" process. Defendants' will use their CAL based platform to identify and prioritize all substantively similar documents over the complete corpus of ESI. The attorneys will then review and code a judgmental

sample of at least 2,500 computer suggested documents to ensure their proper categorization and to further calibrate the system by recoding documents into their proper categories. The initial 2,500 documents shall be produced by or before December 1, 2019.

- (iv) Defendants will have access to the ESI to be searched and will lead the computer training. To ensure transparency and to increase efficiency, Defendants will obtain input from Plaintiff's counsel during the iterative selection and quality control processes, and will share the information used to craft the search protocol as further described herein. All non-privileged documents reviewed by Defendants during each round of the iterative process (i.e. both documents coded as relevant and irrelevant) will be produced to Plaintiff's counsel during the iterative seed set selection process. Plaintiff's counsel will review the documents produced and shall promptly provide defense counsel with its own evaluation of the initial coding applied to the documents, including identification of any documents it believes were incorrectly coded. To the extent the Parties disagree regarding the coding of a particular document, they will meet and confer in an effort to resolve the dispute prior to contacting the Court for resolution. Again, the irrelevant documents so produced shall be promptly returned after review and analysis by Plaintiff's counsel and/or resolution of any disputes by the Court.
- (v) The accuracy of the search processes, both the systems' functions and the attorney judgments to train the computer, will be tested and quality controlled

by both judgmental and statistical sampling. In the statistical sample, a small set of documents will be randomly selected from the total corpus of the documents to be tested. The small set will then be reviewed and an error rate calculated therefrom. The error rates can be reliably projected on the total corpus, having a margin of error directly related to the sample size.

Step 2: Analysis of Initial Random Sample

- (i) Using their identified and disclosed EDiscovery platform, Defendants will generate a random sample of the entire corpus of documents uploaded in their initial search. In order to develop a baseline for calculating recall and precision, Defendants' attorneys will conduct a review of the random sample. To the extent applicable, any relevant documents also will be coded with one or more of the issue tags referenced in paragraph D(3)(d) above. The random sample should consist of 2,500 documents, which represents a 95% confidence level/High Recall rate with a confidence estimation of plus or minus 2%. The Parties agree to utilize the random sample generated prior to the finalization of this protocol. However, during Plaintiffs' counsel's review of the random sample, they may advise as to whether they believe any of the documents should be coded with one or more of the subsequently added issue codes (i.e. Complaints/HR and Defendants/Jurisdiction) and will, as discussed above indicate any disagreement with Defendants' classifications.
- (ii) Judgmental Sampling: In addition to the above, a number of targeted searches where be conducted by Defendants in an effort to locate documents responsive to

several of Plaintiff's specific discovery requests. Approximately X documents have already been coded as responsive and produced to Plaintiff. In addition, several judgmental searches were conducted which resulted in approximately X documents initially being coded as responsive and X number of additional documents have been coded as irrelevant. The documents coded as relevant and non-privileged will also be reviewed by Plaintiff's counsel and, subject to their feedback; they will also be included in the control. An explanation shall be provided by Defendants' attorneys for the basis of the bulk tagging of irrelevant documents (e.g. primarily electronic periodicals and newsletters that were excluded in the same manner as spam junk mail etc.). The explanation shall include the types of documents bulk tagged as irrelevant as well as the process used to identify those types of documents and other similar documents that were bulk tagged as irrelevant.

Step IV: Initial and iterative Training

- (i) Following the creation of the first control set, the Defendants' Software will review the entire data set to identify other potentially relevant documents. Defendants will then review and tag a judgmental based sample, consisting of a minimum of 1,000 documents, including all documents ranked as highly relevant or hot, of the new "Computer Suggested" documents, which were suggested by the software (at a High Recall Rate of at least 90%). Defendants' attorneys shall make a reasonably good faith effort to select documents in the judgmental sample that will serve to enhance and increase the accuracy of their culling.

Step V: Sampling the Excluded Documents

Prior to confirmation of the control or sample set, the Responding Party shall review a Statistically Valid Sample from Excluded Documents, disclose the size of that sample set, and produce to the Requesting Party any responsive, non-privileged documents the Responding Party identifies.

Step 6: Sample Set Review

- (i) The Responding Party shall determine the prevalence of responsive information from the Document Universe by reviewing a Statistically Valid Sample of documents from the Document Universe (the “Sample Set”). Prior to the commencement of Control Set Identification, the Responding Party shall disclose the results of the review of the Sample Set to the Requesting Party, including the number of documents that were coded for each of the Responsiveness Categories during the review of the Sample Set, together with how each document was coded for each of the Responsiveness Categories.

Step 7: Protocol and Time Frames for Raising Disputes

- (i) The Requesting Party shall raise any disputes regarding how the documents were coded for each of the Responsiveness Categories in the Control/Sample Set within five (5) business days of the production of 2,500 documents or less, or within ten (10) business days of the production of more than 4,000 documents. The parties agree to meet and confer in good faith over any such disputes. All non-responsive documents produced from the Control/Sample Set shall be deemed “Highly Confidential” under the terms of the Rule 502(d) Order. The non-responsive documents shall be used only for the purpose of evaluating the

accuracy of the document coding, and shall be promptly returned or destroyed after review by the Requesting party and the resolution of any disputes.

- (ii) The Responding Party shall produce all non-privileged documents identified as non-responsive during the training process. The Requesting party shall raise any disputes regarding how the documents were coded within ten (10) business days of their production, and the parties shall agree to meet and confer in good faith over any such disputes. All non-responsive documents produced: shall be deemed “Highly Confidential” under the terms of the Rule 502(d) Order. In the event the Parties deem it necessary, any and all non-responsive documents shall be used only for the purpose of evaluating the accuracy of the document coding; and shall be promptly returned or destroyed after review by the Requesting Party and the resolution of any disputes.
- (iii) Uncategorized Documents: The Responding Party shall disclose the number of documents the Predictive Coding Software is unable to categorize and they shall review such documents in their entirety in order to identify responsive, non-privileged documents for production.
- (iv) Validation Set: Prior to production, the Responding Party shall review a Statistically Valid Sample of documents in the Document Universe that are categorized as non-responsive by the Predictive Coding Software (the “purported Non-Responsive Documents”) in order to determine the prevalence of responsive documents that are contained therein. Prior to production, the Responding Party shall disclose to the Requesting party in writing the number of

Purported Non-Responsive Documents, the size of the Validation Set, the number of documents identified as responsive during the review of the Validation Set, and the implied rate of Recall. The Responding Party shall produce any responsive non-privileged documents it identifies during the review of the Validation Set.

E. Sources

The Parties have identified the following sources of potentially discoverable ESI. Phase I sources will be addressed first, and Phase II sources will be addressed after Phase I source searches are complete. Sources marked as “N/A” will not be searched by the Parties.

	Data Source	Description	Phase
a	Outlook Web Access	Hosted on HHC’s server at the host address: nychhc.org	1
b	Epic Remote Access	Located at epicremotedesktop.org	2
c	GroupWise E-Mail	Legacy corporate system that provided e-mail communication and calendaring functions	1
d	Home Directories	Personal network storage locations on file server(s)	1
e	Shared Folders	Shared network storage locations on the file server that are accessible by individual users, groups of users or entire departments.	1
f	Database Servers	Backend databases (e.g. Oracle, SQL, MySQL) used to store information for front end applications and for other purposes	2
g	PeopleSoft	Human resources information management system	1
h	PRISM	PeopleSoft component used for time and billing	1
i	Desktops/Laptops	Fixed and portable computers provided to employees to perform work related activities.	1
J	KRONOS	Provides information on Employee’s time and leave usage and timesheets.	1
k	BMC		

l	Mobile Communication Devices	Portable PDAs, smart phones, tablets used for communications	1
m	Removable Storage Devices	Portable storage media, external hard drives, thumb drives, etc. used to store copies of work related ESI	1

F. Custodians

1. The Parties seek the Court's intervention for the designation of the custodians and time periods. The Chart below captures the parties' proposals for both:

	Custodian's Name	Plaintiff's Proposed Time Period	Defendants' Proposed Time Period
1.	Elizabeth Ford	January 1, 2011 to present	same
2.	Abhishek Jain	April 2018 to present	same
3.	Ross MacDonald	January 1, 2015 to present	January 1, 2017 to present
4.	Jonathan Wangel	January 2018 (start date with H & H) to present	Same
5.	Patricia *Patsy" Yang	January 1, 2015 to present	December 1, 2015 to present
6.	Mitchell Katz	March 2018 to present	Objects to this custodian
7.	Melissa Kaye	January 1, 2011 to present	January 1, 2011 to present
8.	Andrea Swenson	January 1, 2018 to present	Same
9.	Anansa Brayton	January 1, 2018 to present	Same

G. Format of Production of Documents

1. PDF/Native File Format Production: Documents will be produced as single page PDF files with corresponding multi-page text, necessary load files and bates stamped accordingly.

The load files will include an image load file as well as metadata (.DAT) file with the metadata fields identified on Exhibit X. Defendants will produce spreadsheets (.xls files) and PowerPoint presentations (.ppt. files) in native form as well as any documents that cannot be converted to Pdf format (e.g. audio, or video files, such as mp3s, wavs, megs, etc.). In addition for any redacted documents that are produced, the documents' metadata fields will be redacted where required. For the production of ESI from non-email sources, the parties will meet and confer to attempt to reach an agreement of the format of production. In addition to any paper production, Defendants shall produce all responsive documents in .pdf and native format on CD-ROM.

2. Appearance: Subject to appropriate redaction, each document's electronic image will convey the same information and image as the original document. Documents that present imaging or formatting problems will be promptly identified, and the parties will meet and confer.
3. Document Numbering: Each page of a produced document will have a legible unique page identifier, "Bates Number" electronically "burned" onto the image at a location that does not obscure, conceal, or interfere with any information from the source document. The Bates Number for each page of each document will be created so as to identify the producing party and the document number. In the case of materials redacted in accordance with applicable law or confidential materials contemplated in any Confidentiality Stipulation entered into by the parties, a designation may be "burned" onto the document's image at a location that does not obscure or conceal any information from the source document.

4. Production Media: The producing party will produce documents on readily accessible, computer or electronic media as the parties may hereafter agree upon, including CD-ROM, DVD, external hard drive (with standard PC compatible interface), (the “Production Media”). Each piece of Production Media will be assigned a production number or other unique identifying label. Such label shall contain the date of the production of documents on the Production Media (e.g. Defendant City Production December 1, 2019), as well as the sequence of the material in that production (e.g. D001-002) For example, if the production comprises document images on three DVDs, the producing party may label each DVD in the following manner “Defendant Production December 1, 2019; Defendant Production December 1, 2019-002; Defendant Production December 1, 2019-003. Additional information that will be identified on the physical Production Media includes: 1) text referencing that it was produced in: Kaye v. H +H Corp., et. al.; and 2) the Bates Number range of the materials contained on the Production Media. Further any replacement Production Media will cross reference the original Production Media and clearly identify that it is a replacement; it shall also cross reference the Bates Number range that is being replaced.
5. Write Protection and Preservation: All computer media that is capable of write-protection should be write-protected before production.
6. Inadvertent Disclosures: The terms of the Parties’ Claw back Agreement and Court Order shall apply to this protocol.
7. Duplicate Production Not Required: A party producing data in electronic form need not produce the same document in paper format.

H. Timing

1. To the extent a timeframe is not specifically outlined herein, the parties will use their reasonable efforts to produce ESI in a timely manner consistent with the Court's discovery schedule.
2. The parties will produce ESI on a rolling basis.

I. ESI Preservation

1. Defendants have issued litigation hold/legal hold/preservation notices to designated employees and stakeholder on the following dates (PLEASE PROVIDE)

J. General Provisions

1. Any practice or procedure set forth herein may be amended by agreement of the parties, and will first be confirmed in writing, where such variance is deemed appropriate to facilitate the timely and economical exchange of electronic data.
2. Should any party subsequently determine it cannot in good faith proceed as required by this protocol, the parties will meet and confer to resolve any dispute before seeking Court intervention.
3. The Parties agree that e-discovery will be conducted in phases and, at the conclusion of the search process described herein, the Parties will meet and confer regarding whether further searches of additional custodians and/or the Phase II sources is warranted and/or reasonable. If agreement cannot be reached, either party may seek relief from the Court.
4. The Parties hereby agree to ***meet and confer in good faith*** over any disputes that might arise with respect to the terms and conditions of this Order or any

other aspects relating to discovery. The Responding Party agrees to make its Subject Matter Expert reasonably available to the Requesting Party for questions about its use of predictive coding. Should the Parties be unable to resolve their disputes on any issues stemming from the use of predictive coding as set forth in this Order, they shall submit those issues to the Court for resolution.

5. Notwithstanding the provisions set forth in this Order, the Parties respectively reserve their rights regarding the instant discovery process. This includes, but is not limited to, the Requesting Party's right to object to the efforts of the Responding Party to search for, review, and produce information in response to the Requesting party's document requests, and the Responding Party's right to withhold information pursuant to the objections it previously interposed in response to the Requesting party's document requests.

IT IS SO ORDERED:

DATE:

HON. J. PAUL OETKEN
United States District Judge

Special Hagan Esq.,
Law Offices of Special Hagan
196-04 Hollis Avenue
Saint Albans, New York 11412
(917) 337-2439
special@haganlawoffices.net
Attorney for Plaintiff, Melissa Kaye, M.D.

James Johnson
Corporation Counsel of the City of New York
Attorney for Defendants
100 Church Street,
New York, New York 10007
(212) 356-2461
dcanfield@law.nyc.gov